Rising Star Evaluation in Heterogeneous Social Network

Feng Ding, Yuqing Liu, Xin Chen, Feng Chen,

Abstract—Rising stars are junior individuals in the social network who will have high impacts with time accumulation. Rising star evaluation has become a research hotspot in network analysis area recently, which is helpful for decision support, resource allocation, and other practical problems. As a traditional social network, academic social network is stressed because of its heterogeneity and regular data structure. In this paper, we assume there are inside factors influencing individuals behaviors. We process the network parameters and mine inner factors via factor analysis, and train a decision tree to evaluate furture impact. Experiment is processed on America Physics Society(APS) dataset, and the result shows our method has better performance than state-of-the-arts.

Index Terms—Rising Star, Factor Analysis, Decision Tree, Hetegeneous Social Network

1 Introduction

Rising star evaluation is an issue of network impact analysis, which is highly considered in wireless sensor network(WSN), ad-hoc and other dynamic networks[1]. Rising stars are nodes with high potential of impact. When newly joining a circle, they may have little impact. With time accumulation, they will turn to be senior ones quickly. Rising star evaluation is a prediction of network impact[2]. It is not so helpful to evaluate the impact of new ones because of the few actions, while prediction will reveal the impact in future[3].

Works of network analysis are based on factors of nodes[4]. There are 2 factors influencing the individuals' future impact. On the one hand, a rising star must be potential, which is the inner factor. On the other hand, position of node in network shows how it will communicate with others. Nodes at the center of network are easier to accumulate impact, while those at edge of the network are harder to show their impact. Assume nodes are same, then the only factor is the position. There are algorithms aiming at position and structure analysis, which perform well in homogeneous networks.

However, it is different in social network situations. Nodes in social network turn to be heterogeneous with more properties. The properties of social nodes are meaningful, which cannot be considered similar simply. As a traditional social network, academic social network is much more considered because of its practical meaning and reliable dataset. Academic social network is one kind of heterogeneous network. There are 3 kinds of nodes in academic social network: author, paper, and journal. Among nodes there are links with the meaning of co-author, citation and publishment. The complexity of academic social network shows the difficulty of evaluation and prediction.

There are rising stars in social network, especially in academic social network. There are junior scholars newly joining academic social network with few papers and citations. With time accumulation, rising stars of them will become senior ones and gain large citations quickly. Citation is one important indicator to evaluate a scholar's success. We regard it as the index of impact.

Finding academic rising stars is a meaningful job. Most research institutions prefer hiring employees with rich experience. In fact, more job-seekers are freshmen. So how to make the better choice is a key problem for institutions. Rising star evaluation is also helpful for journal editors. When choosing peer-reviewers, editors want to invite highly regarded reviewers. Sometimes these senior scholars are too busy to review the paper, so rising stars come to be the best substitutes. Meanwhile, inviting rising stars more is helpful for them to become experienced. So it is a win-win way. For postgraduates, needless to say the importance of choosing mentor. Students hope to learn from senior scholars. In fact, many students are led by junior one in the name of senior's. So telling rising stars will be helpful for the supervisor choose. Similarly, it is also meaningful for researchers to choose cooperators.

There are methods for finding rising stars in social network. The basic algorithm is the PageRank[5] proposed by Page L et al. in 1999. PageRank is used to calculate the importance of web-pages, which shows the impact of nodes. The main idea of PageRank is random walk of weights among the entire network. PageRank consider the linking of nodes. If node A links to node B, then the weight should flow from A to B. PageRank considers the linking relation as flowing direction, which is opposite to academic social network. PubRank[6] proposed by Xiao-Li Li et al. is an improved method for finding rising stars in social network. Different from PageRank, PubRank convert the weight flow

- *F. Ding and Y. Liu are with the School of Software, Dalian University of Technology, Dalian, China.*
- *F. Chen is with the Software Technology Research Laboratory, De Montfort University.*
- *X. Chen is with the School of Electronics and Information Engineering, Liaoning University of Technology, Jinzhou, China; Corresponding author: chenxin.lut@hotmail.com*

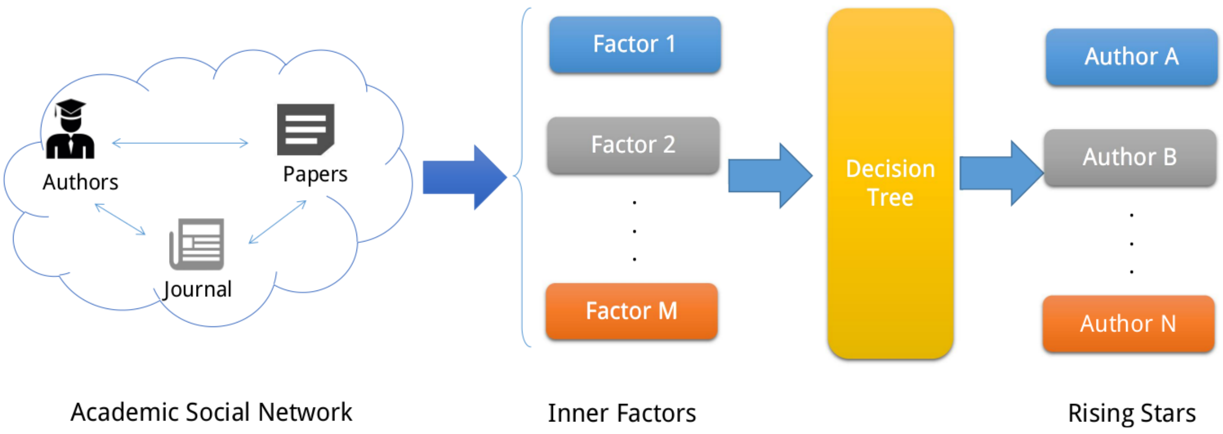

Fig. 1. The Structure of Our Method

direction. Scholars learn from other authors and papers, which decide the direction is opposite to PageRank. Based on PubRank, A Duad et al. come up with StarRank[7] method to evaluate rising stars with improved parameters. Although the performance is reliable, those methods are time consuming because of iteration.

Some impact evaluation methods are based on network parameters instead, which have high speeds. H-index[8] is a scientific index widely used for scholar analysis. It considers both paper count and citations, showing an overall ability of an author. As a supplement, G-index[9] is highly correlative with H-index, which is normally higher in numeral. Although G-index is more friendly, it is hard for evaluating rising stars because of the only paper-citation based parameters. Indeed, such indexes show the accumulations of research, which are not trustworthy for impact prediction, that is, rising star evaluation.

How can we combine the speed and performance? The advantage of PageRank-based methods is considering the network structure. If there are structure parameters correlative with the impact, we could try to put forward a function or model to make evaluation like H-index. Indeed, there are relations among network parameters and inner factors of nodes[10]. From the work of Wang D et al.[11], success of scholars is decided by fortune and research quotient. Suppose fortune is equal for everyone, then the research quotient counts. Success is revealed in behaviors shown as network parameters. From this hypothesis, we try to predicate the inner factors from explicit network parameters, and apply evaluation method to the judgment. To speed the computation, we should process data dimension reduction[12]. In our work, factor analysis is an ideal model for dimension reduction.

Decision tree is a typical model for decision support with tree topology[13]. As a traditional tool in machine learning, decision tree is built considering properties of all situations. After training this model to minimize the entropy of data selection, we can classify samples or make regression. There are 2 kinds of decision tree: regression and classification. In our work, we train a regression tree to make predictions of scholar impacts and finding rising stars. The framework of our work is shown in Fig. 1.

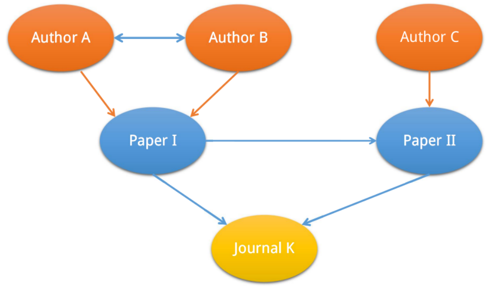

Fig. 2. An Simplified Demo of Academic Social Network

Structure of this paper is as following. Firstly we introduce our work in chapter one. Then we will define the problem and introduce related work separately in chapter two and three. Our method is specifically introduced in chapter four. The experiment result shows our method has better performance and less time cost than the baseline in chapter five.

2 Problem Definition

Before presenting our work, some definitions for the research problem should be proposed. In this section, we give the concrete concept of academic social network and propose the formular presentation. Meanwhile, we define the social impact in academic network, apply citation to standing for the impact, and give the reason.

2.1 Academic Social Network

Formally, an academic social network can be defined as $G =< V, E >$, where V is the set of vertexes in G, and E is the set of relations among V. Typically, since academic social network is heterogeneous, there are 3 kinds of vertexes in V: author, paper and journal. Different kind of nodes has disparate properties. A simplified network is shown in Fig. 2.

From the figure, there are practical relations. Link between author and paper shows the relationship that he or she composes the paper. If there are authors composing a same paper, then links among writers show the cooperations. Usually, one paper is supposed to cite others' works. Links between papers shows the citations. Journals publish the papers, which lead to the link between papers and journals. All links have directions. If an author links to another one, then we consider that they have the cooperation relationship. Obviously, this is a double-way link. Link between author and paper has one direction, which means that the author writes this paper. If authors link to same paper, then they must connect each other. Links among papers show the relations of citations, so they are one-way links. If paper P_I links to paper P_{II}, then we consider P_I cites P_{II}. From the concept of citation, there is no circle in links among papers. Finally, link between paper and journal shows the relation that paper P_I is published on journal J_K.

Besides links, there are properties of nodes shown in the network structure. Formally, a node can be described a unique tuple of neighbors and links. In social network, a node N_a can be described as

$$N_a =< Sn_a, Se_a > \tag{1}$$

where $Sn_a = \{N_k | N_k \textit{ is linking to or linked by } N_a\}$, Se_a is the set of links.

All properties of nodes are processed by data from Sn and Se. In academic social network, an author's appendix properties include his or her paper number, cited times, and etc. All the properties are calculated from the number or weight of special links.

2.2 Social Impact

After building the network, we are supposed to calculate the social impact. Social impact is an abstract concept with different practical meanings in different situations. From the work of Bell S.M. et al.[14], social impact means the ability of influencing and changing others in social network. Similarly, impact in academic social network means a scholar's ability of influencing others. If one scholar has higher social impact, others will learn from him or her. Higher social impact means higher authority.

Social impact in academic social network is mainly shown as citations and cooperations. If author A_a cooperates with author A_b, then A_a and A_b will learn from and affect each other. More times A_a and A_b cooperate, more they will affect each other. If one scholar has higher social impact, he or she must has more cooperators. This conforms to common sense. Citation is another way for scholar to study other's work. If paper P_I cites paper P_{II}, then we consider P_I's work learns from P_{II}. The authors of P_I also learn from authors of P_{II}. From this fact we can find that if a scholar has higher social impact, his or her papers will be cited more times. This corresponds to the facts.

There are indicators to evaluate social impact. Considering cooperation, if a scholar has more cooperators, then the scholar's impact will be higher. There are lacks to regard cooperation as the only indicator. In fact, one scholar has limit energy to enlarge the circle. Meanwhile, if there is a circle of junior scholars, it is hard to say the impact is high no matter how large the cooperator number is.

In this paper, citation is considered as the appearance of social impact. A scholar with higher impact will be followed by more researchers, which accelerates his or her citations. From this fact, we hold the hypothesis that citation number will present the social impact index. There are facts supporting our hypothesis. Yann LeCun, one of the most famous scholars of maching learning, is cited 62665 times. The book of Hawking[15] holds the citation number 11333. (All of the data are from Google Scholar, Dec. 27th, 2017). Obviously, if a scholar is famous, then he or she will be cited more times. His or her citation number per article will also be large. In this paper, we regard the total citation number as

$$Citation_a = \sum_{p \in P_a} Cited(p) \tag{2}$$

where P_a is the paper set that published by author A.

3 Related Work

In this section, we will introduce some methods about finding rising stars. Considering the lack of H-index and G-index, most works about rising star evaluation are based on network structure and random walk. The typical one is PageRank[5] proposed by Page L. et al. Based on PageRank there are PubRank[6], StarRank[7], WMIRank[16], and etc. We will mainly introduce PageRank and PubRank.

3.1 PageRank

PageRank is one of most famous network analysis methods. The most considered factor in PageRank is the link. PageRank believes if a node has high impact, then it will be linked more times. In other words, an outstanding node will have more edges linked to it. If node A holds higher authority, then other nodes with high authority will link to A. Considering there is an authority flow. If node A linked by others, then others' authority will flow to this node. Meanwhile, node A will transfer its authority by linking to others. With the flowing of authority, there is a time the network converging. This is the core of PageRank.

PageRank is calculated as following:

$$PR(A) = \frac{1-d}{N} + d(\sum_{n \in N_A} \frac{PR(n)}{L_n}) \tag{3}$$

where N is the number of nodes, N_A is the node set that linking to A, L_n is the number of edges that linking from n.

3.2 PubRank

PageRank turns to be success in Internet situation, which is homogeneous. Considering heterogeneous network, there are different parameters in the evaluation. PubRank is a method to evaluate bibliography network based on PageRank. Different from PageRank, PubRank considers heterogeneous parameters and extend the weight of authority flow.

Instead of separate authority averagely, PubRank weights the flow by the paper's academic level and the strength of cooperations.

PubRank defines the publication quality score of an author, which is used to show the ability of research. Publication quality score is defined as following:

$$\lambda(A) = \frac{1}{|P_A|} \sum_{p \in P_A} \frac{1}{\alpha^{pub(p)-1}} \tag{4}$$

where P_A is the paper set of A, $pub(p)$ is the publication quality score of paper p, and α is a constant.

Finally, PubRank is calculated as following:

$$PUB(A) = \frac{1-d}{N} + d * \sum_{n \in N_A} \frac{w(A,n) * \lambda(A) * PUB(n)}{\sum_{i \in N_n} w(i,n) * \lambda(i)} \tag{5}$$

where $w(A, n)$ is the cooperation strength between A and n.

PubRank regards the publication quality of papers as the quality of published journals, for the paper published in short time has few citations. In our experiment, we try to replace it with the paper's PageRank and show the results.

4 Method

In this section, we will introduce our method to find rising stars. From the hypothesis, we believe there are inner factors influencing the outside behaviors and impact. The task of evaluation is to dig out them from outer network parameter and evaluate future impact with them. We will introduce each step specifically as following.

4.1 Network Parameter Extraction

To explore the inner factors of social behaviors, the first step is to confirm outer features that cover them as many as possible. Outer features of one scholar are shown in network as structure parameters. There are 3 kinds of nodes in the heterogeneous academic social network. Edges show the relationship. Both edges and nodes describe the structure of network, which represent parameters of nodes. These parameters describe outer features of a scholar.

Network parameters chosen for analysis should cover different types of nodes and relations as many as possible. On the one hand, it is better to describe a specific node with the only parameter vector and its neighbors. On the other hand, too many parameters will lead to high computation complexity. From the definition of academic social network, links between author and other nodes contain publishing, co-author, citing and cited relations. So the parameters chosen should focus on them.

Define first-step relation as parameters of nodes linking to or linked from authors. It is easy to find that first-step relation covers the relations of cooperation, publishment, and citation. First-step relation is a local descriptor, which cannot show the impact transmission. So we consider second-step relation, which contains the first-step relations of author's neighbors. Suppose there are average N neighbors of a node. Then the count of first-step relation is $O(N)$, and $O(N^2)$ for second-step, which is too large. So we should prune the relation. We only consider the citation in second-step relations.

To consider the first-step relation, we choose 9 parameters covering the activity and authority. To show the activity, we consider the paper number of an author, the co-author number, the active year, the citing times, the citing times per paper, and the citing times per year into consideration. Co-author number shows the communication of the author, and parameters about citing times shows the understanding of state-of-the-arts. The active year means the years that a scholar publish papers. To show the authority, we consider the cited times, the cited times per paper and the cited times per year.

To consider the second-step relation, we choose 3 parameters. They are the citied time of papers one author cites, the citing time of papers one author cites, and the cited time of co-authors.

Meanwhile, we bring triple closure into parameter analysis. Triple closure is one of important concept in network science. In this situation, relations among nodes are separated into strong ones and weak ones. Strong relation means 2 nodes have more communications. Weak relation means they are not close friends but acquaintances. Triple closure claims if node A has strong relaton with node B, node B has strong relation with node C, then A has at least weak relation with C. If it holds, then we say A follows the principle of triple closure.

The triple closue in academic social network is shown in Fig. 3. Author A has cooperated with author B, and author B has cooperated with author C. If A follows the triple closure principle, then A will cooperate with C. The situation of triple closure shows whether A could learn more from others, and the impact of A's work. If A does not follows the principle, which means A cannot learn from C. Meanwhile, C cannot learn A's work, and the impact of A is restricted. We define the parameter of strong co-authors as following:

$$StrongCoau(A) = \sum_{i \in N_A} \sum_{j \in N_i} isneb(j, A) \tag{6}$$

where

$$isneb(j, A) = \begin{cases} 1 & j \in N_A \\ 0 & else \end{cases}$$

and N_A is the co-author set of A.

From the definition, if an author's strong co-author value is higher, then he or she has more triple closures, which means his or her communication circle is tighter. With tighter circle, the author will learn from others more.

4.2 Factor Analysis

From our hypothesis, there are inner factors influencing authors' outer behaviors. After processing network parameters, we should dig out the inner factors. In this paper, we make the hypothesis that inner factors are linearly dependent with network parameters. On the one hand,

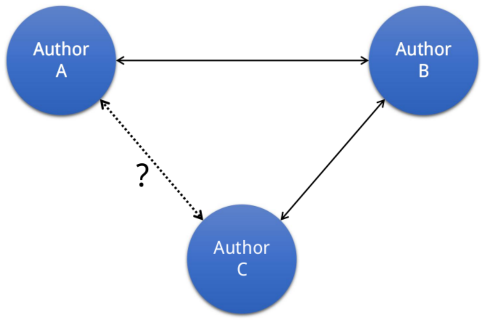

Fig. 3. An Example of Triple Closure

linearly dependence is easy to fit and reduces the computation complexity. On the other hand, from the experiment result, we can find that linearly dependence leads to good performance.

To solve this problem, we apply factor analysis. Let a_i be the i-th sample of author list. Formularly, we can describe a_i as $a_i = [p_{(i,1)}, ..., p_{(i,n)}]$. $p_{(i,k)}$ means the k-th processed parameter of a_i. Let vector $f_i = [f_{(i,1)}, ..., f_{(i,m)}]$ be the inner factor of a_i. From the hypothesis, there is:

$$a_i = \mu_i + X f_i + \delta_i \tag{7}$$

where μ_i is the bias of a_i, X is the weight matrix, and δ is the noise. Re-write Eq. 7 in matrix formular, we get

$$A = \mu + XF + \Delta \tag{8}$$

where F is the matrix of factors, Δ is the matrix of noise. From the factor analysis, Δ is considered Gaussian distributed with mean 0 and convariance Ψ, i.e. $\Delta \sim N(0, \Psi)$. From the definition of Δ, there is $cov(F, \Delta) = 0$. As X is normalized, we have $E(F) = 0$ and $Var(F) = I$. From the definition, $\Psi = diag(\psi_1, ..., \psi_m)$. Specially, if $\Psi = I$, then this method becomes traditional PCA.

Now, our task is to determine the variables. To optimize these matrixes, we choose EM method for computation. Finally, we get the weight and bias matrixes, which can be used to extract factors of new samples.

4.3 Decision Tree

After factor analysis, we should evaluate the impact with these inner factors. Decision tree is traditional machine learning model for classification and regression. From our hypothesis, inner factors influencing the outer behaviors, which decide the impact of an author. In other words, we evaluate the impact with the decision of inner factors. So decision tree is a suitable model for impact evaluation.

Impact evaluation is a regression problem. We regress the impact from inner factors. In this paper, we choose CART as the training model. Classification and regression trees(CART) is a binary decision tree for classification and regression. The structure of a trained CART is as Fig. 4.

From Fig. 4, yellow nodes are statuses, and blue nodes are result. The links between yellow and blue nodes shows the decision. For example, if a sample is in status 1, and it

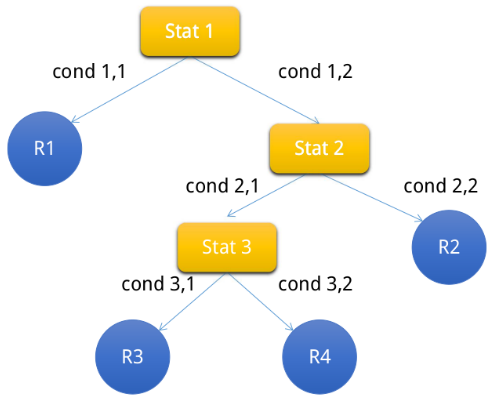

Fig. 4. Structure of CART

fits the condition 1,1, then this sample belongs to result 1. If not fitting the condition, it will come into status 2 and be judged continuously until the sample comes to a leaf node.

To make regression, the leaf nodes are some simple regression models rather than specific labels. The conditions for decision are learned from the gini impurity. Gini impurity is an index to evaluate the condition's performance. The definition of gene impurity is as following:

$$G(f) = \sum_{i=1}^{n} f_i(1 - f_i) = 1 - \sum_{i=1}^{n} f_i^2 \tag{9}$$

where f is the separation condition, f_i is the chosen probability of kind i under the condition f. The task of building condition nodes is to minimize the gini impurity.

5 Experiment

5.1 Experiment Environment

Our experiments are processed on APS dataset. America Physics Society dataset(APS) is one of typical datasets, which including the paper and citation information. There are works processing on APS dataset. Different from DBLP dataset, APS dataset contains the citation relation. APS dataset are composed as json files, which can help us build the heterogeneous social network easily.

In our experiment, we process 4 sub-datasets from 1970 to 2010, separated by 10 years. In each sub-dataset, we process the 13 network parameters for analysis. For each 20-year time, we choose the first 10 years' dataset as train data and regard the next 10 years' data as label. For example, we process the social gene from 1970 to 1980, and use the rank from 1980 to 1990 as label. The model trained is used to evaluate rising stars from 1980 to 1990. Different intervals' experiments are similar.

We choose PubRank as comparison. Considering the lack of author's publication quality score in PubRank, we use PageRank as the quality of paper instead, and call it extend-PubRank. We train out decision tree with labels as PubRank and extend-PubRank separately. In our experiment, Decision-Pub means training with label of PubRank, and Decision-Ex means training with extend-PubRank. Since we choose 10 years as an interval, all rising stars begin their

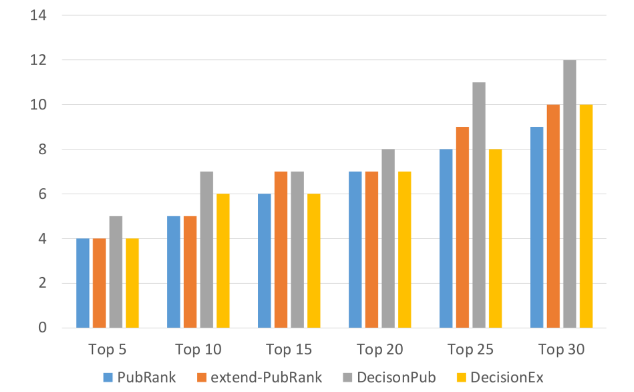

Fig. 5. Hitting Count of Top Rising Stars from 1970 to 1980

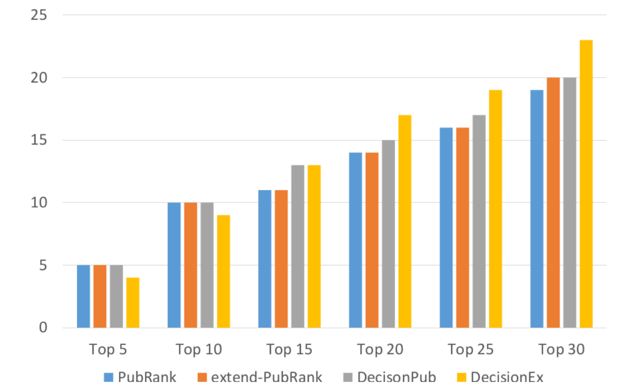

Fig. 6. Hitting Count of Top Rising Stars from 1980 to 1990

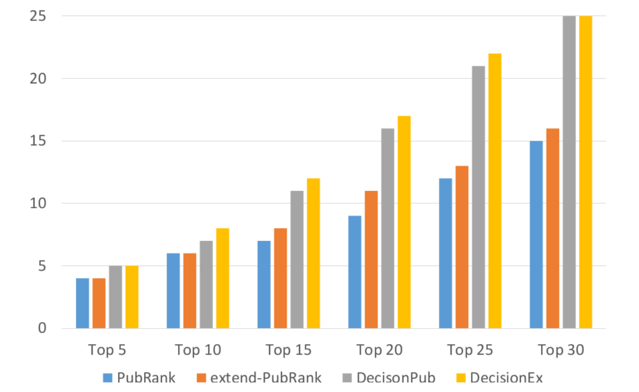

Fig. 7. Hitting Count of Top Rising Stars from 1990 to 2000

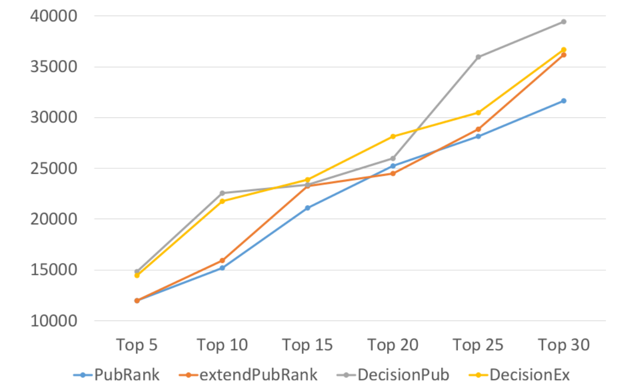

Fig. 8. Citation Count of Rising Stars from 1970 to 1980

careers from the beginning year, and they are active at least over 2 sub-datasets. The results are shown as following.

5.2 Result

To show the performance of outstanding scholar prediction, we list the top 5 rising stars predicted by our method and PubRank in Tab. 1, 2, and 3.

From the table, we can find that our methods are better than the comparisons in most situations. The best scholar of top 5 evaluated by PubRank method is clearly evaluated by our method. Meanwhile, our methods predict 2 rising stars with citation higher than 10000 during the years from 1980 to 1990, which PubRank does not find out.

To qualify the performance of top author's prediction and show the accuracy of methods, we choose the rate of hitting top authors as index for analysis. In this part, we regard authors who have top 15% citation number in rising star dataset as the top authors. The results are shown in Fig. 5, 6, and 7.

From the results, we can find that our methods can hit more top authors than PubRank. Specially in the period from 1990 to 2000, our methods have reached the accuracy more than 80%.

To show the performance in a macro view, we calculate the sum of citation within 3 periods, which shows an accumulated impact of methods. The results are shown in Fig. 8, 9, and 10.

From the result, we can find that our methods hold higher citation than PubRank. Different period of years holds different citation count. With time accumulation, the difference will be enlarged, which shows our methods have a better performance than PubRank.

6 Conclusion

In this paper, we try to evaluate the rising stars via inner factors. Firstly, we process the outer behaviors and apply factor analysis to mining the inner factors. Then we train a decision tree with PubRank to find the rising stars. Result shows our method holds better performance than the comparison method.

In the future, we will try to train decision tree with different ranks, and test the method on more datasets. Meanwhile, we find that the performance on dataset from 1990 to 2000 is much better than the performance from 1970 to 1980. We will try to find the reason.

References

[1] Zhang Q, Lin M, Yang L T, et al. Energy-Efficient Scheduling for Real-Time Systems Based on Deep Q-Learning Model. *IEEE Transactions on Sustainable Computing*, 2017, PP(99):1-1.

[2] Ning Z, Liu Y, Kong X. Social gene A new method to find rising stars. *International Symposium on Networks, Computers and Communications*, IEEE, 2017:1-6.

PubRank		extend-PubRank		Decision-Pub		Decision-Ex	
Author	Citation	Author	Citation	Author	Citation	Author	Citation
G. J. Feldman	4239	G. J. Feldman	4239	G. J. Feldman	4239	E. I. Shibata	4730
K. W. Kemper	1587	K. W. Kemper	1587	E. I. Shibata	4730	G. J. Feldman	4239
J. D. Garrett	1522	J. D. Garrett	1522	E. W. Plummer	3131	M. Suenaga	1241
H. R. Weller	706	Howard Georgi	3922	C. S. Wang	1495	E. W. Plummer	3131
Howard Georgi	3922	H. R. Weller	706	M. Suenaga	1241	C. Stassis	1113

TABLE 1
Top 5 Rising Stars Evaluated from 1970 to 1980

PubRank		extend-PubRank		Decision-Pub		Decision-Ex	
Author	Citation	Author	Citation	Author	Citation	Author	Citation
R. J. Cava	6053	R. J. Cava	6053	P. Sinervo	10998	A. A. Volkov	2001
Ivan K. Schuller	2879	Ivan K. Schuller	2879	S. Errede	11509	M. Marx	1282
T. Murakami	1886	R. Poling	6741	I. Yu	9503	P. Sinervo	10998
R. Poling	6741	R. Kass	9355	F. Cervelli	4180	S. Errede	11509
R. Kass	9355	T. Murakami	1886	S. Cihangir	7253	I. Yu	9503

TABLE 2
Top 5 Rising Stars Evaluated from 1980 to 1990

PubRank		extend-PubRank		Decision-Pub		Decision-Ex	
Author	Citation	Author	Citation	Author	Citation	Author	Citation
J. Zhang	11697	P. Fallon	2232	A. H. Castro Neto	5385	J. Zhang	11697
P. Fallon	2232	J. Zhang	11697	Daniel Loss	7065	A. H. Castro Neto	5385
Andrew R. Liddle	2890	Andrew R. Liddle	2890	M. Garcia-Sciveres	5986	Daniel Loss	7065
M. Lewitowicz	979	Z. V. Vardeny	1099	L. Zhang	9044	L. Zhang	9044
M. Wang	2154	K. Honscheid	10221	J. Zhang	11697	W. C. Wester, III	8125

TABLE 3
Top 5 Rising Stars Evaluated from 1990 to 2000

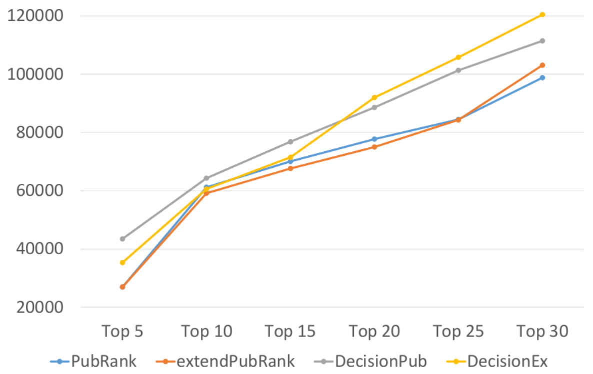

Fig. 9. Citation Count of Rising Stars from 1980 to 1990

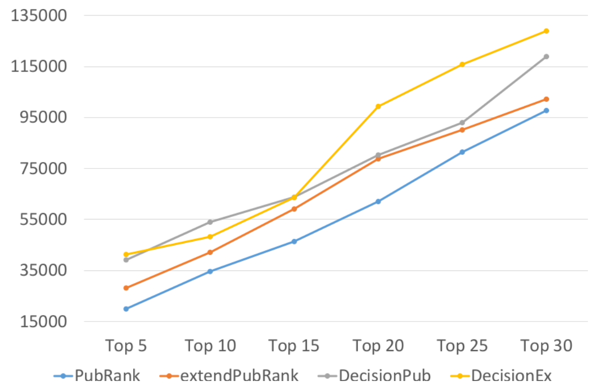

Fig. 10. Citation Count of Rising Stars from 1990 to 2000

[3] Zhang Q, Chen Z. A Distributed Weighted Possibilistic c-Means Algorithm for Clustering Incomplete Big Sensor Data. *International Journal of Distributed Sensor Networks*, 2014, 2014(2):4-4.

[4] Zhang Q, Yang L T, Chen Z, et al. Privacy-preserving Double-projection Deep Computation Model with Crowdsourcing on Cloud for Big Data Feature Learning. *IEEE Internet of Things Journal*, 2017, PP(99):1-1.

[5] Page L, Brin S, Motwani R, et al. The PageRank citation ranking: Bringing order to the web. Stanford InfoLab, 1999.

[6] Li X L, Foo C S, Tew K L, et al. Searching for rising stars in bibliography networks. *International Conference on Database Systems for Advanced Applications*, Springer Berlin Heidelberg, 2009: 288-292.

[7] Daud A, Abbasi R, Muhammad F. Finding rising stars in social networks. *International Conference on Database Systems for Advanced Applications*, Springer, Berlin, Heidelberg, 2013: 13-24.

[8] Hirsch J E. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(46):16569-16572.

[9] Egghe L. Theory and practise of the g-index. *Scientometrics*, 2006, 69(1):131-152.

[10] Zhang Q, Yang L T, Chen Z, et al. PPHOPCM: Privacy-preserving High-order Possibilistic c-Means Algorithm for Big Data Clustering with Cloud Computing. *IEEE Transactions on Big Data*, 2017, PP(99):1-1.

[11] Sinatra R, Wang D, Deville P, et al. Quantifying the evolution of individual scientific impact. *Science*, 2016, 354(6312):aaf5239-aaf5239.

[12] Zhang Q, Yang L T, Chen Z, et al. An Improved Deep Computation Model Based on Canonical Polyadic Decomposition. *IEEE Transactions on Systems Man & Cybernetics Systems*, 2017, PP(99):1-10.

[13] Zhang Q, Yang L T, Chen Z, et al. A survey on deep learning for big data. *Information Fusion*, 2018, 42:146-157.

[14] Bain S K, Bell S M. Social self-concept, social attributions, and peer relationships in fourth, fifth, and sixth graders who are gifted compared to high achievers. *Gifted Child Quarterly*, 2004, 48(3): 167-178.
[15] Hawking S W, Ellis G F R. The large scale structure of space-time. Cambridge university press, 1973.
[16] Daud A, Aljohani N R, Abbasi R A, et al. Finding Rising Stars in Co-Author Networks via Weighted Mutual Influence. *International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 2017:33-41.

Feng Ding Feng Ding received his B.E. degree in Power Engineering from Dalian University of Technology, China, in 2002 and Ph. D degree in Management Science and Engineering from Dalian University of Technology, China, in 2009. He is currently an associate professor in School of software Technology in Dalian University of Technology, China. His research interests include e-government, e-commerce, decision support, and information security.

Yuqing Liu Yuqing Liu received his B.E. degree in Software Engineering from Dalian University of Technology, China, in 2017. He is currently working toward M.Sc. degree at Dalian University of Technology. His current research interests include data mining, operation research, and image processing.

Xin Chen Xin Chen received his B.E. degree in computer science from Dalian University of Technology, China, in 2005 and Ph. D degree in information science from Poznan University of Technology, Poland, in 2014. He is currently an associate professor of school of electronic and information engineering in Liaoning University of Technology, China. His research interests include combination optimization, especially scheduling problem and algorithm design.

Feng Chen Feng Chen received BSc, M-phil and PhD at Nankai University, Dalian University of Technology and De Montfort University in 1991, 1994 and 2007. He is a senior lecturer at De Montfort University. His research interests include software engineering, knowledge engineering, Pervasive computing, Ambient Assisted Living (AAL) and image processing.